Dustin Doose, Lee Halbritter, and Madison Winker
STAT 442
2015-2019 Salary Analysis in Connecticut
Dec 5, 2019


Dealing with large data sets can be a daunting task, especially when that data requires quite a bit of manipulation and clean up.  Four years of salary data from the state of Connecticut is an example of a large data set in need to cleaning.  The goal of this endeavor is to explore salary data and rise in pay as it pertains to males and females across various categories such as age, location, or job type.  Before diving into very specific categories, we'll first take a broad look at the average salaries over the past four years regardless of age, ethnicity, or job.  While this isn't a deeply informative graph, it gives an idea of what we may encounter with just a first glance.
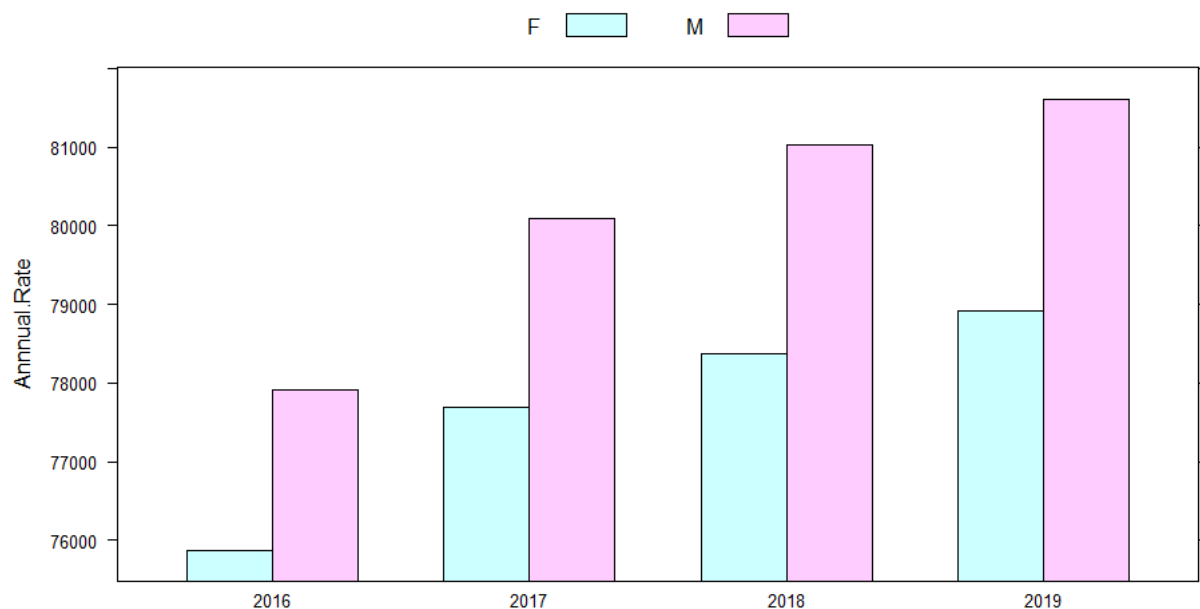


The above plot shows a steady increase in pay for both genders, however the gap in pay between males and females also appears to stay fairly consistent overall.  Let's take a deeper look within some specific categories and see if this trend holds throughout.
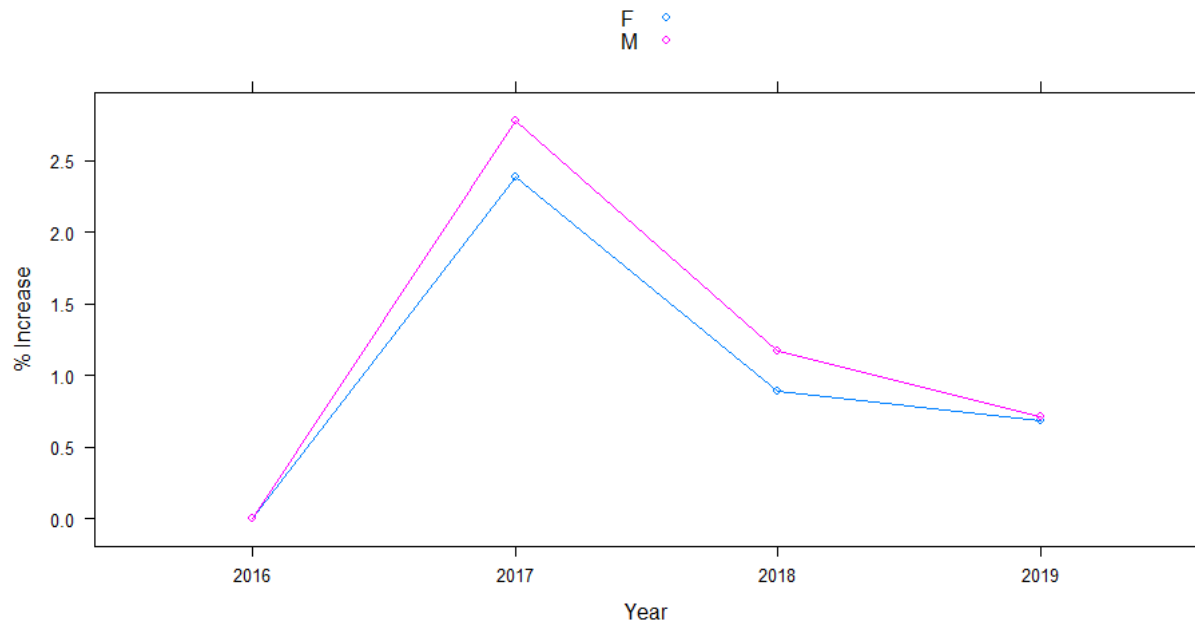
The graph above shows that both genders saw an annual increase in pay, males received a slightly higher increase.  Note the increase shown for 2016 is zero since no prior data was included in the analysis.  Suppose we separate the groups further by ethnicity?  Will the same trends hold or will there be different behavior across the various ethnic groups?
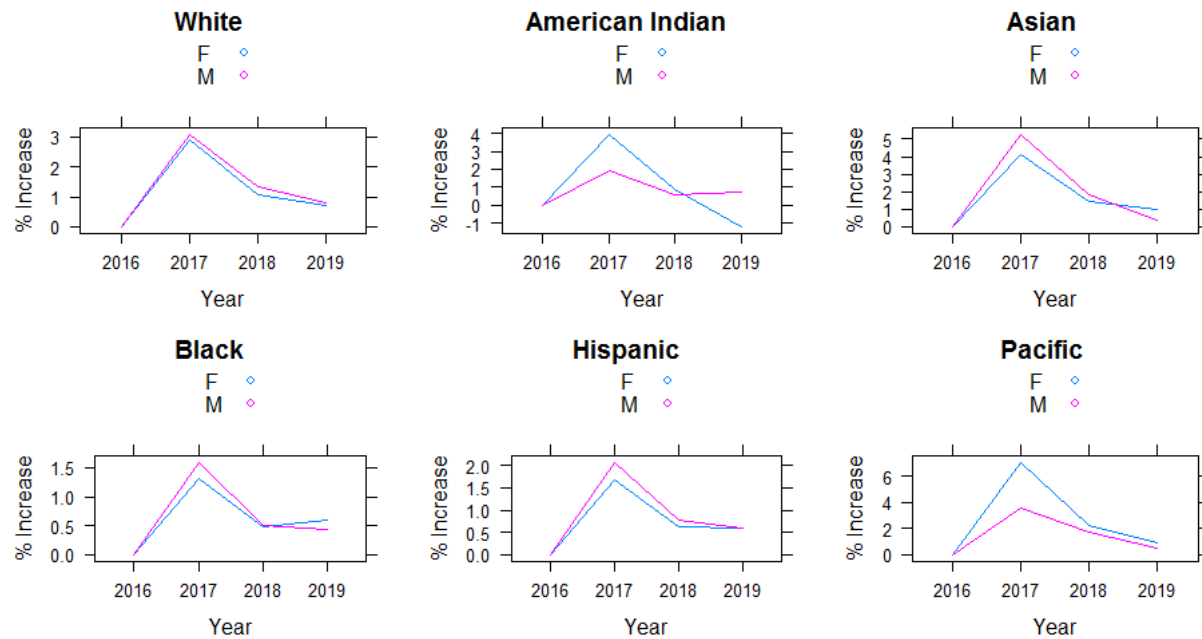
The plots above actually show some different trends. First and foremost, in the American Indian and Pacific groups, females actually received higher increases with the exception of 2019 where female Native Americans actually received a decrease in pay. The other ethnic groups all held a similar trend to what we observed in the overall data.

Let's set ethnicity aside now for the moment and look at some of the most common jobs and how their pay increases were distributed.
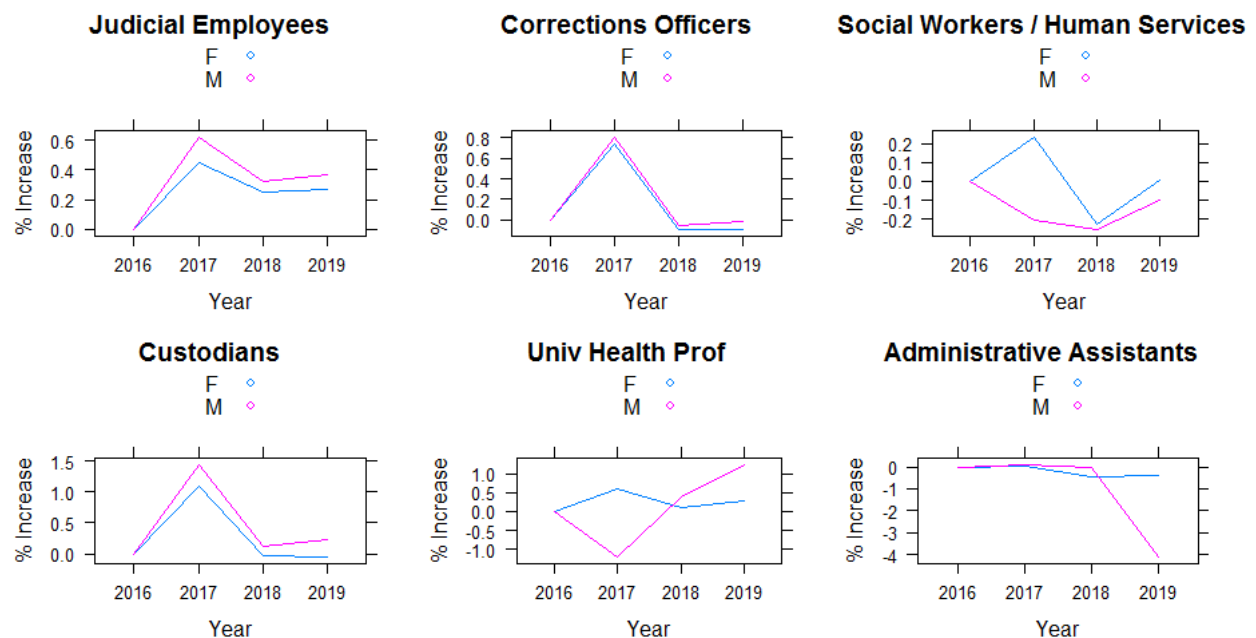


Several of the most frequent occupations in the data were chosen to observe the pay increase trends again. Here we see a variety of results. Judicial Employees, Corrections Officers, and Custodians followed the same general trend as the overall data originally observed. Males received slightly higher annual increases. However, the trends were different for Social/Human Services, University Health Professors, and Administrative Assistants. Females fared better in Social/Human Services, though the field as a whole took a reduction in pay from the 2016 level. Administrative Assistants were fairly flat at around 0% until the whole field took a dip in 2019, males in particular.

Now let's observe the trends when we take age into account.
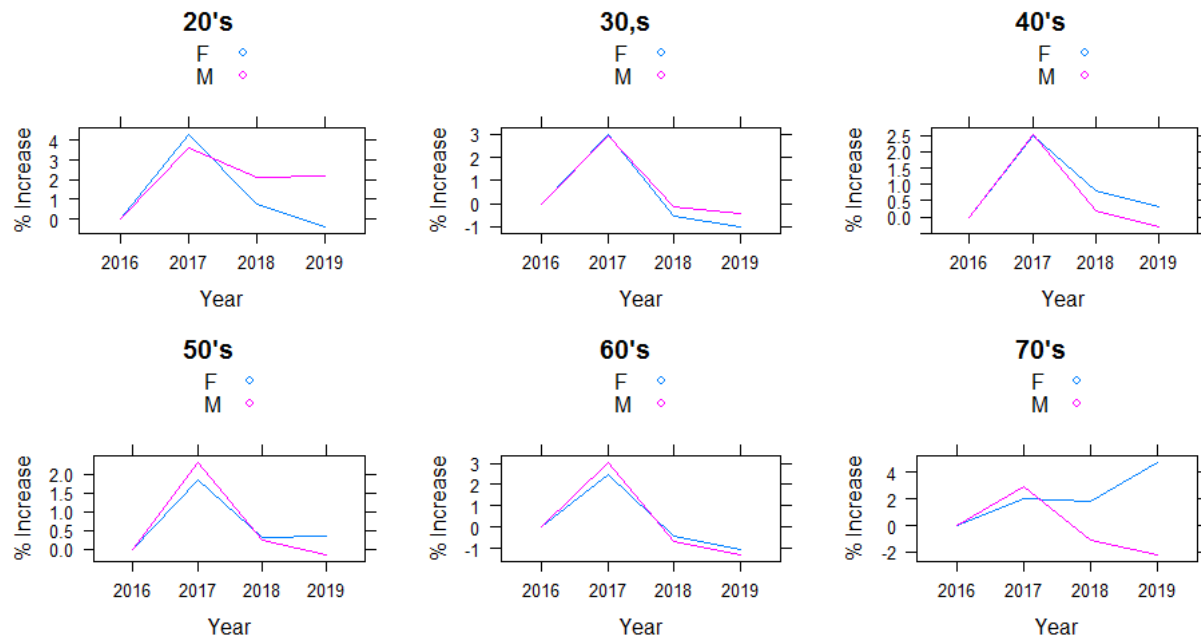


Looking at the age breakdown, it appears that males tend to get larger increases in their 20s and 30s before females gradually overtake them. The discrepancy between the increases is at its largest at the extreme ends of the age range. Over the middle of the age range, the rates are fairly close together.
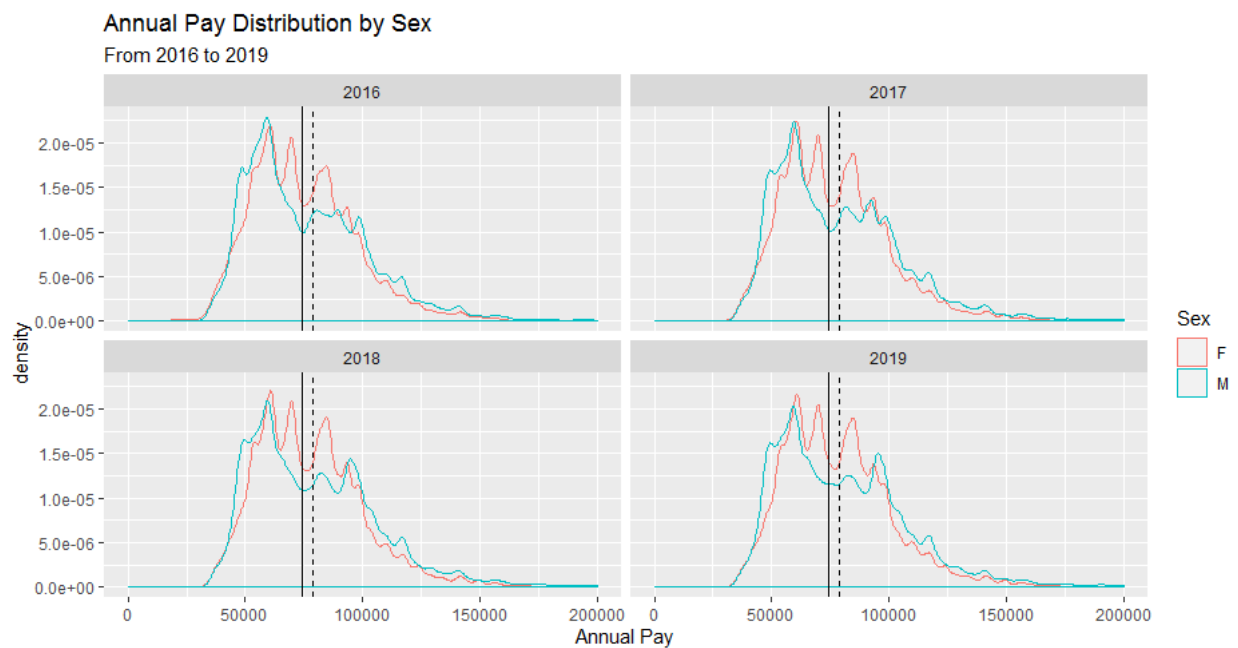
Now looking at the density plots for annual salary for each year, we see that both distributions are skewed slightly to the high side. This is illustrated by the fact that the mean (dashed line) is greater than the median (solid line) in each year. We actually see the male distribution is denser on the low end than the female distribution.

Finally, one thing we notice about most of the previous plots was that 2017 saw a consistently larger pay increase for both genders across many ethinic, age, and job groups. While there are of course exceptions to this, it did appear to be a common theme. Let's see if we can quantify this at all. Did 2017 specifically treat everyone equally well, or did men still make out better?

Analysis of Variance Table
Response: Increase

|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| agegroup | 6 | 5.6569 | 0.94281 | 1.4790 | **0.3233** |
| Sex | 1 | 1.1737 | 1.17367 | 1.8411 | **0.2236** |
| Residuals | 6 | 3.8249 | 0.63748 |  |  |

Analysis of Variance Table
Response: Increase

|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| Ethnic.Grp | 6 | 35.521 | 5.9201 | 4.2894 | **0.04986 *** |
| Sex | 1 | 0.811 | 0.8107 | 0.5874 | **0.47251** |
| Residuals | 6 | 8.281 | 1.3802 |  |  |
| --- |  |  |  |  |  |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Using a one-way ANOVA test, we look at the average pay increase across each age group and each gender. For the one-way ANOVA test, the Null Hypothesis is that the mean of each group is equal. Here we see p-values in excess of 0.2 indicating that the pay increases in 2017 could be assumed to be the same across the board for age or gender.

However, when we look at gender and ethnicity, we do see a borderline p-value for ethnicity indicating that at least one ethnic group differed from the rest. Indeed we can see this visually from the previous plots. Recall in the age group plots, all of the values in 2017 were between 1.5 and 4 percent. However, when we look at the plots for ethnic groups, we see a wider range going from about 1.3 to as much as 6 percent.

In the end, 2017 looked to benefit males and females about equally over all, though there was a greater disparity between different ethnic groups.